# CREDIT EDA CASE STUDY

SANKET BHUWAD

IIITB DS C37

Purpose

- Credit EDA will help the Loan providing company to make decision for Loan Approval based on clients profile.

- It helps to understand the driver variables/factors which are strong indicator of default and use this knowledge for portfolio and risk assessment

- This will control financial loss of the Company

Steps

1. Data Sourcing and Understanding
2. Data Cleaning
   - Data Quality check
   - Handling missing values, Outliers
   - Checking Distribution of Data, Data Imbalance Ratio
   - Performing Univariate and Bivariate Analysis, Correlation Matrix.
3. EDA on Previous Application Data
4. Insights

Distribution of 'TARGET' variable
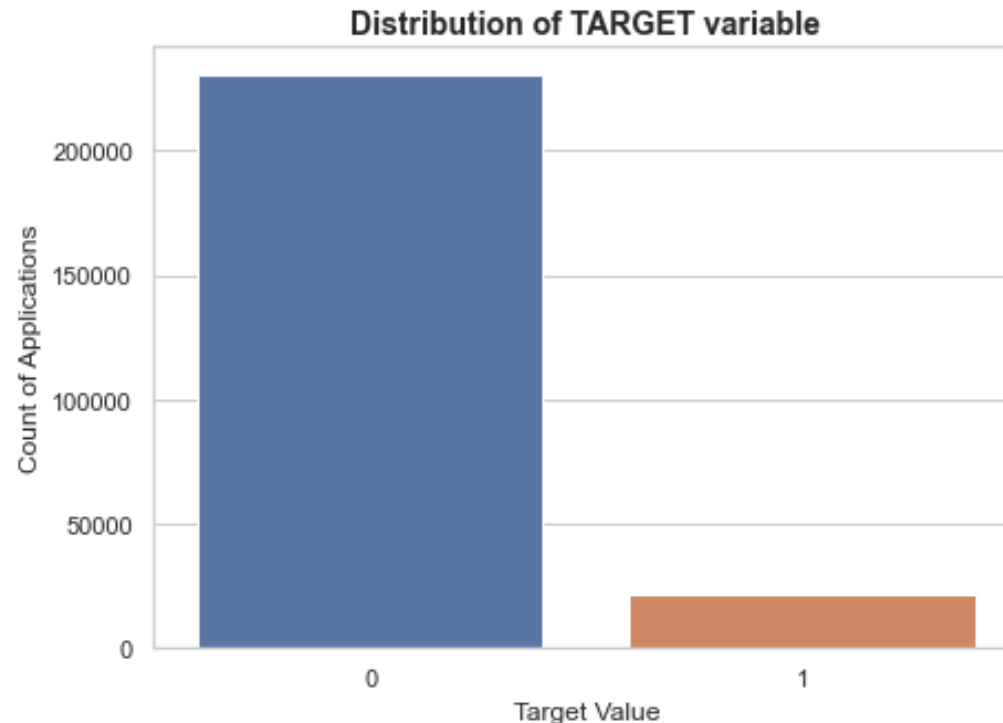
From the plot we conclude that more than 200000 clients are without payment difficulties on the other side only around 22000 are clients with payment difficulties

- The percentage of clients who paid their loan on time is 91.34%
- The percentage of clients who did not paid their loan on time is 8.66%
- Data Imbalance ratio is 10.55 (Data is highly imbalanced)

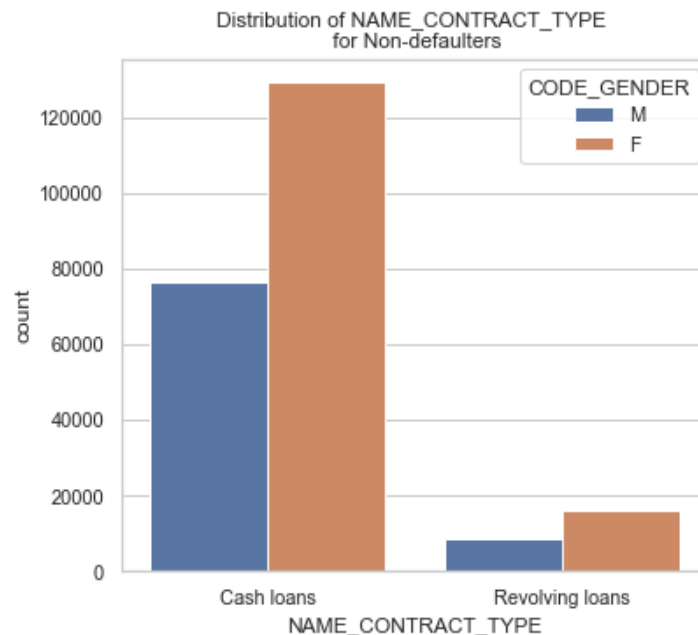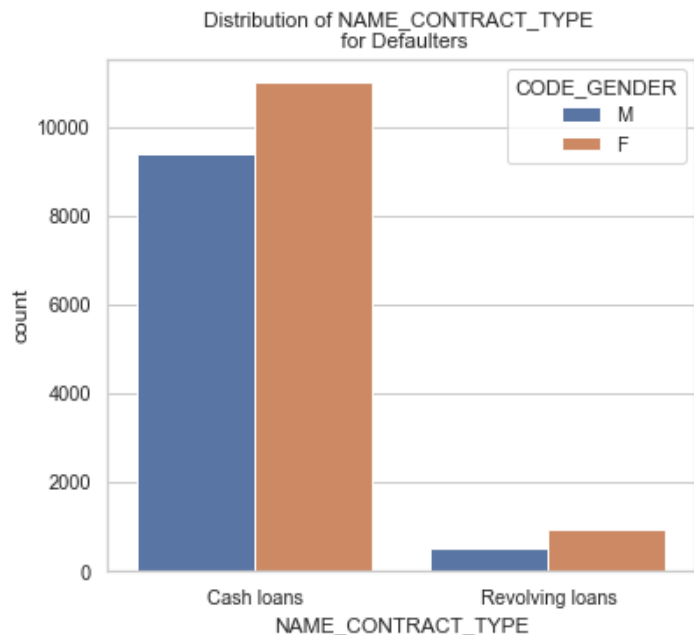# Distribution of clients across Age Group
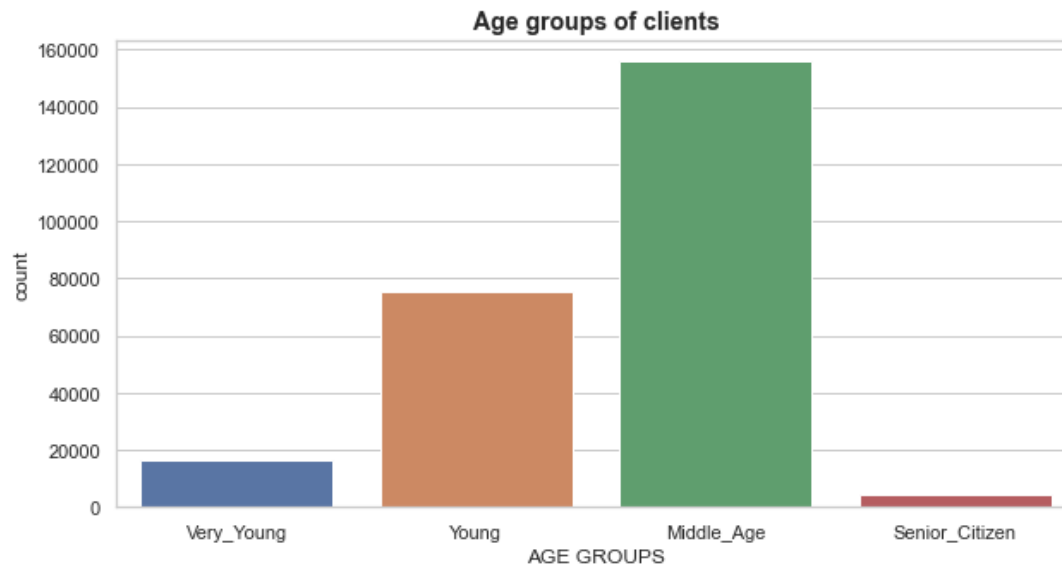
- From this graph we conclude that majority of clients are in range of 25 to 50 i.e. Middle age

Distribution of Income Range

Income range of Clients
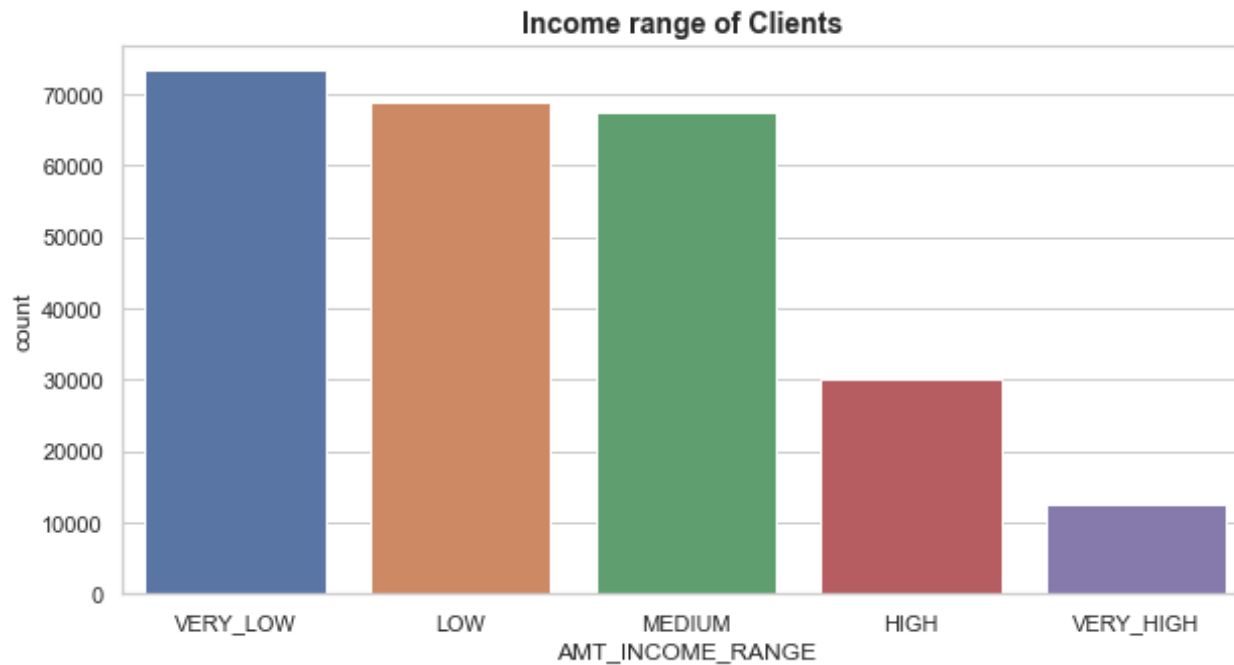
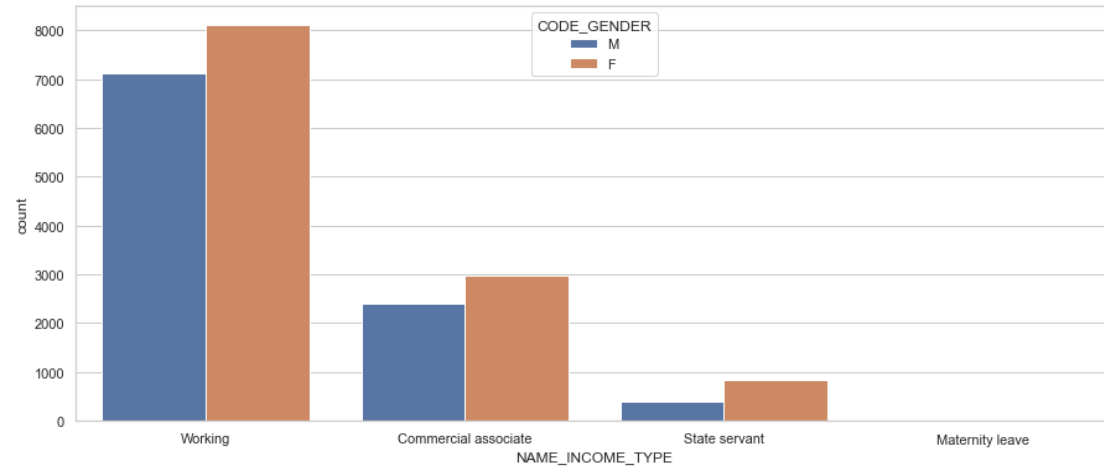• Most of the clients fall in Low Income Range Category
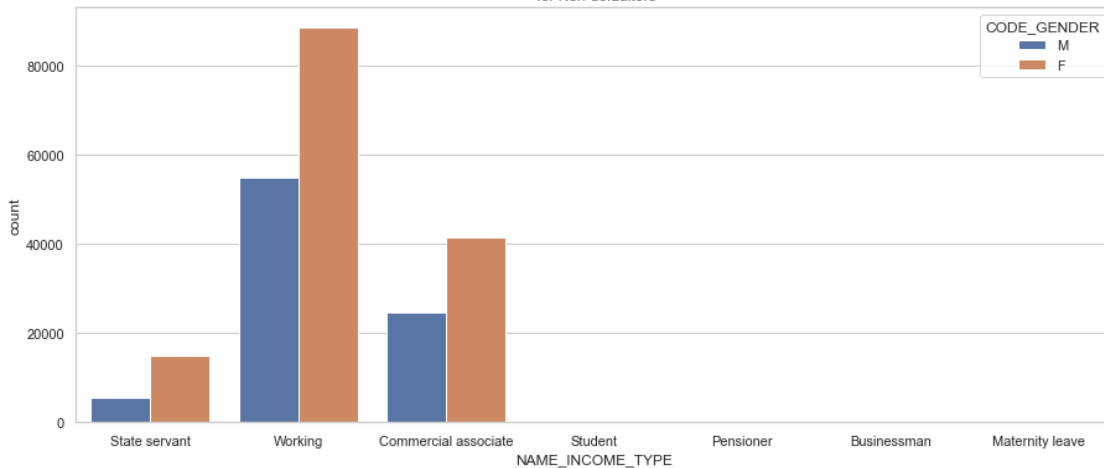
Distribution of Clients across Contract Type

- In both the cases of Defaulters and Non-defaulters Cash loans are preferred rather than revolving loans
- Female clients are leading in both the cases

Distribution of Name Income Type

Distribution of NAME_INCOME_TYPE for Defaulters

Distribution of NAME_INCOME_TYPE for Non-defaulters

- For income type 'working', 'commercial associate', and 'State Servant' the number of clients are higher than other i.e. 'Maternity leave.

- We can see more female clients than male

- Although there are less number of clients for income type 'Maternity leave'.

- In Non-defaulters there is no income type for 'student', 'pensioner' and 'Businessman' which means they don't do any late payments.
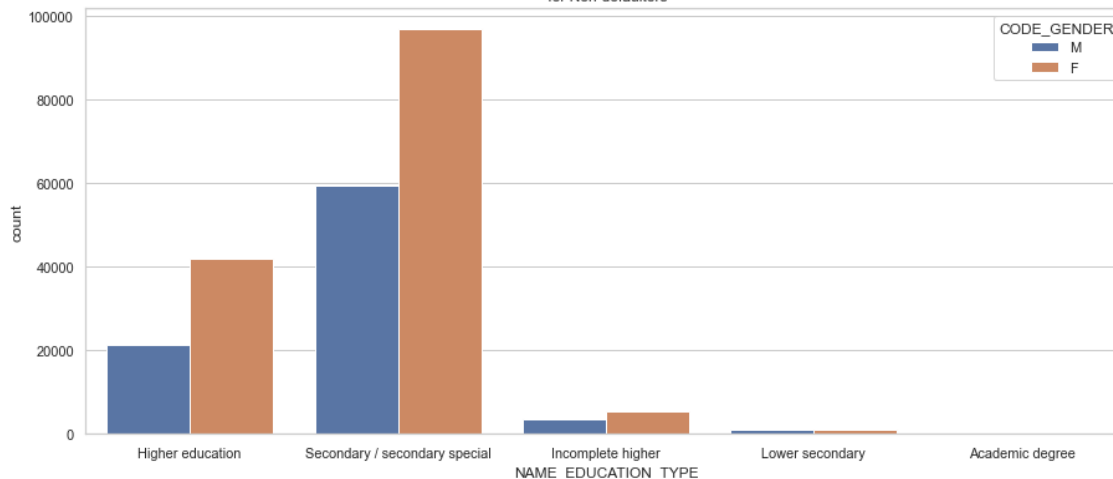
Distribution of Name Education Type

Distribution of NAME_EDUCATION_TYPE for Defaulters

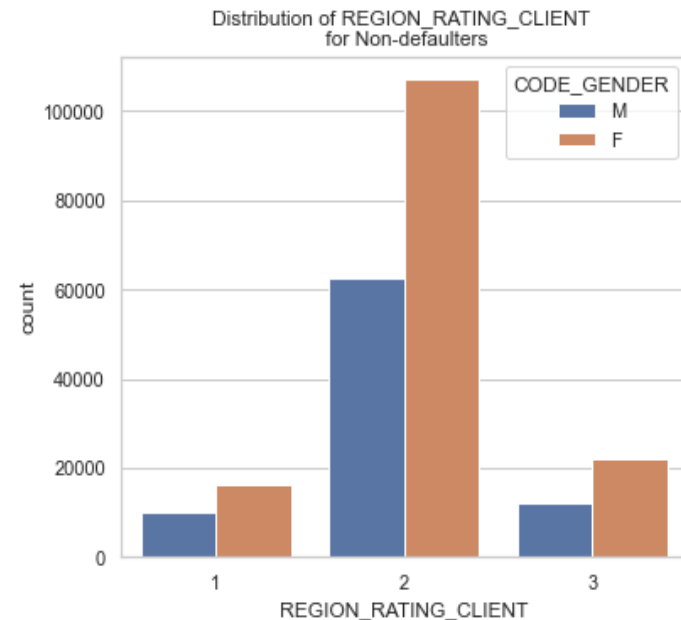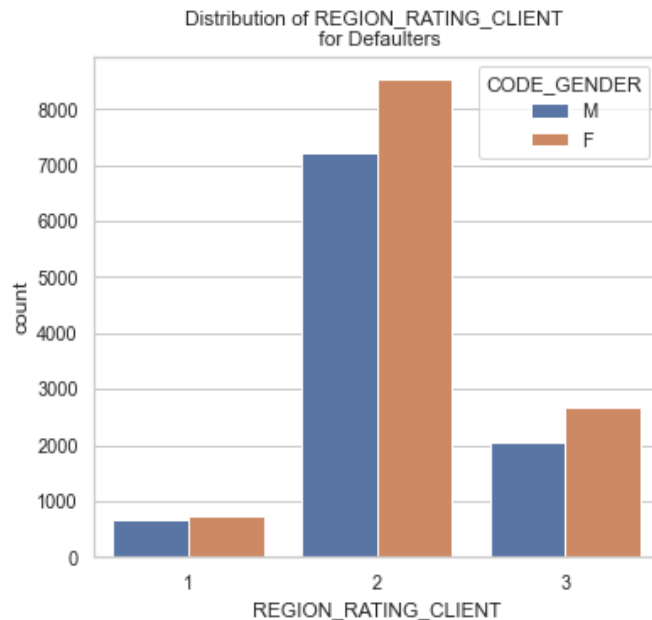Distribution of NAME_EDUCATION_TYPE for Non-defaulters
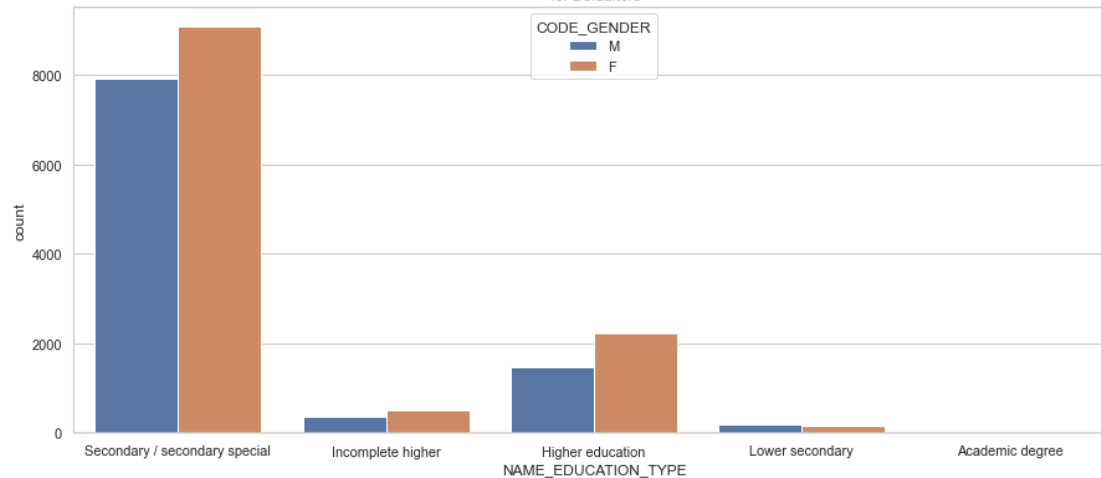
- From the plots we can see clients with higher education and academic degree are less likely to default
- Compare to that secondary education clients are more likely to default
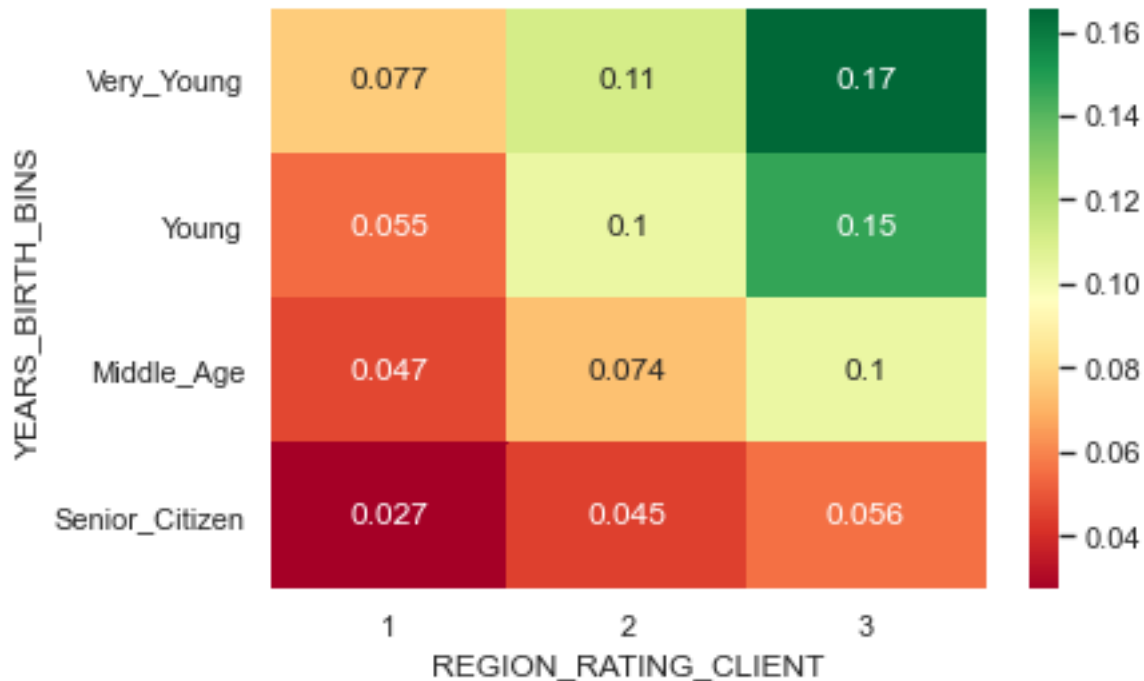
Distribution of Region Rating Client

- Clients from region 2 apply for loan more likely rather than region 1 & region 3
- Selecting clients with highest region rating provided by bank itself will reduce risk clients with highest region rating provided by bank itself will reduce risk of default

Occupation Type vs Name Contract Type Target
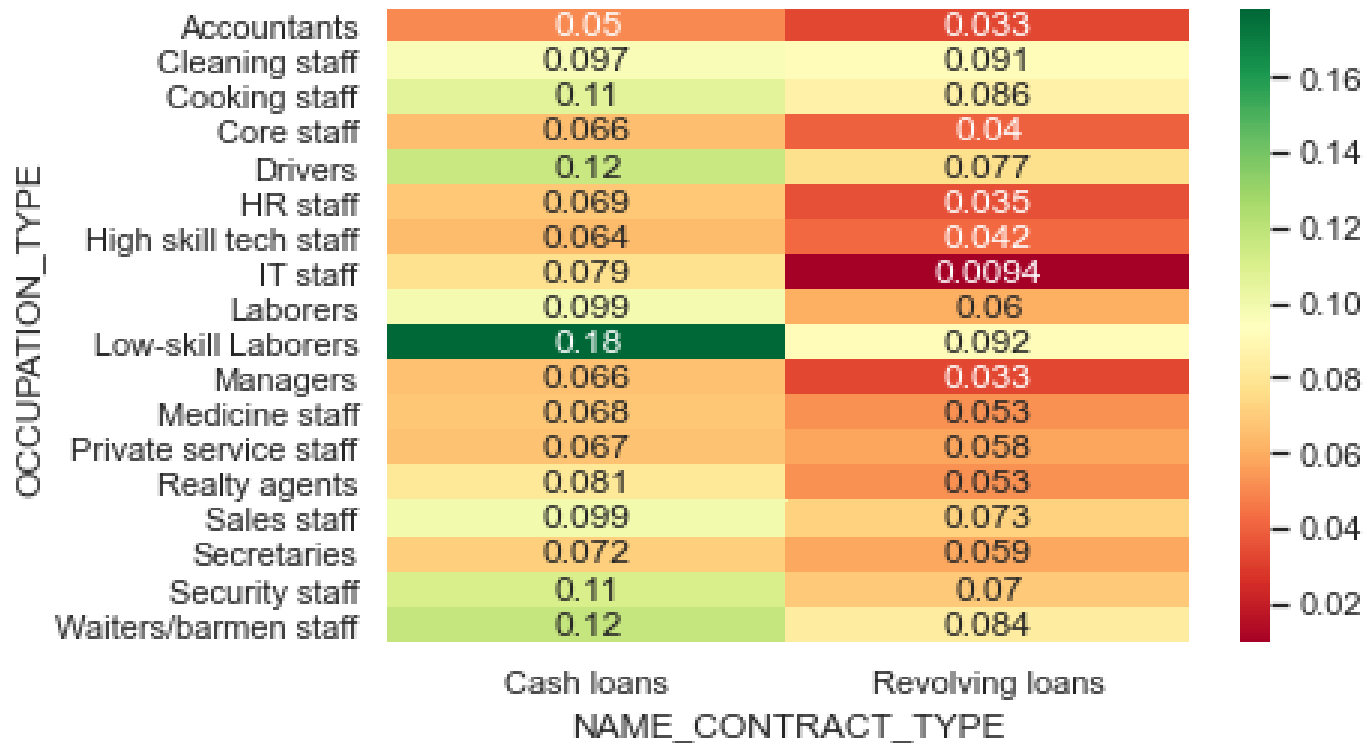

- Clients with OCCUPATION_TYPE like 'Low-skill Laborers','Cleaning Staff', 'Cooking Staff', 'Drivers' are more likely to default

Age Group vs Region Rating Target

- Clients with age group Very Young to Middle Age from region rating 3 are more likely to get default.

- Clients from region rating 1 are less likely to get default irrespective of all age group

Distribution of Contract Status
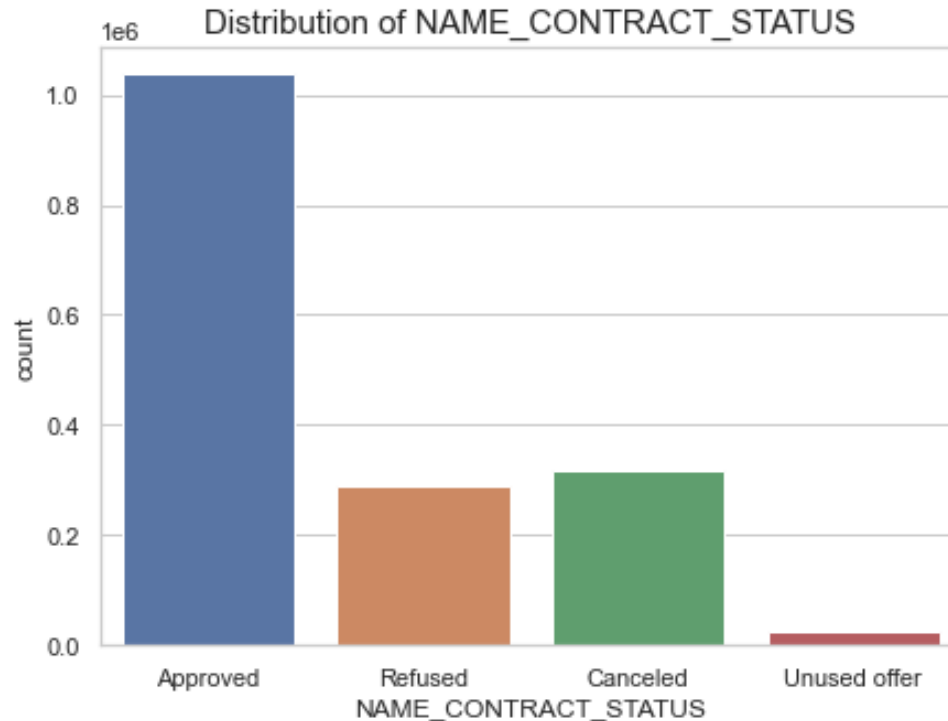

- Checking Previous application data majority of clients application gets approved

Distribution of Name Contract Type


Distribution of NAME_CONTRACT_TYPE
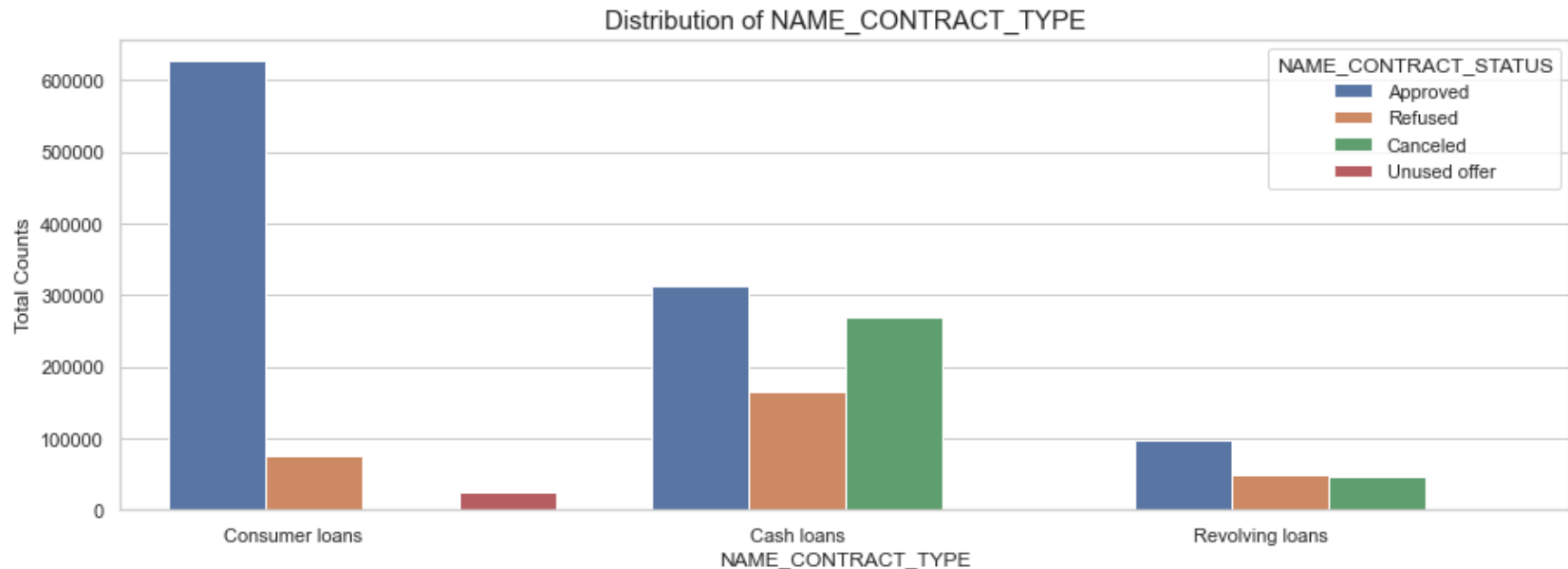
- We can see from previous application data most of the clients apply for 'Cash loans' and 'Consumer Loans'.
- 'Cash loans' are refused compare to 'Consumer loans' and 'Revolving loans'

Income Range vs Previous Contract Status
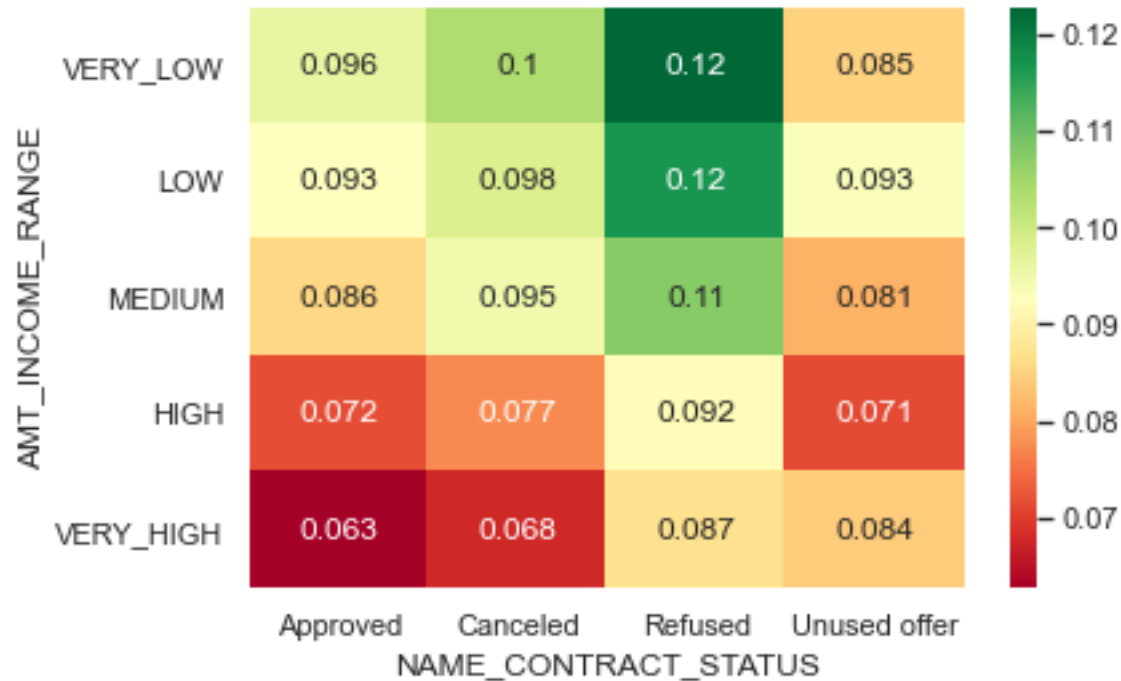
- Clients with very low & low income range are more likely to default whose previous application is cancelled or refused

Housing Type vs Contract Status
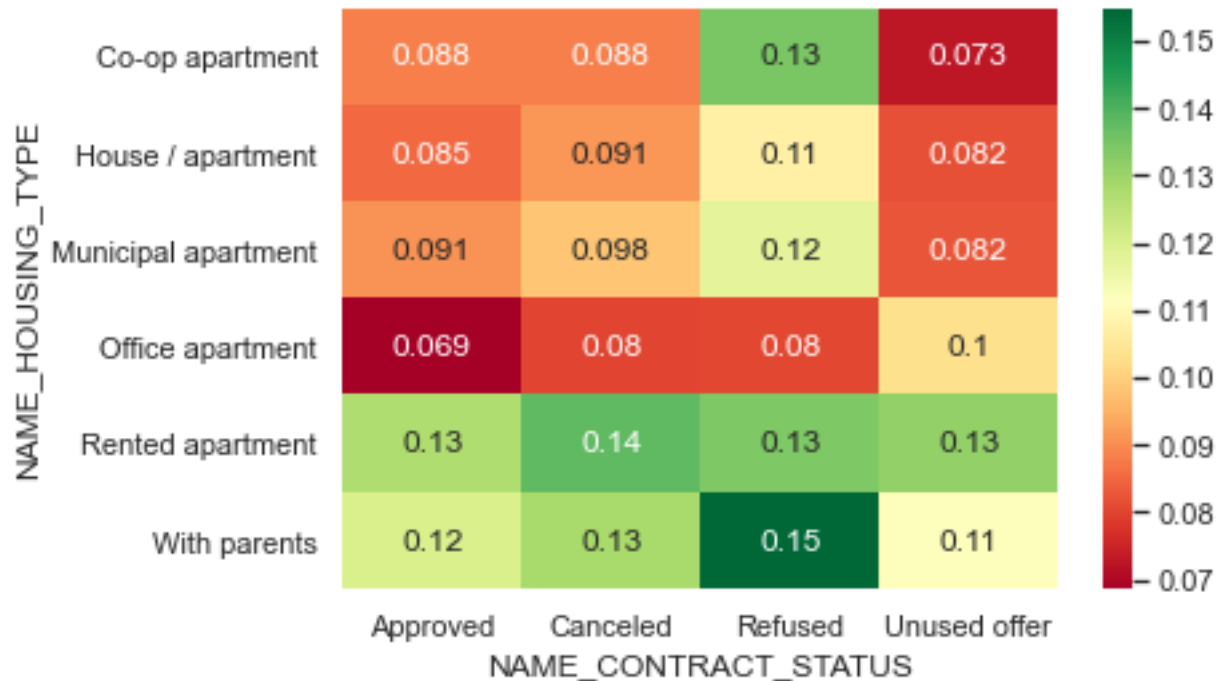
- Clients living in Rented apartments or with parents are more likely to default
- people who live in office apartments whose previous application was approved are less likely to default. company can use this information to reduce their default rate

Recommended clients where loan can be credited

- Senior Citizen with any income group
- Client with high income category
- Clients with higher education
- Any client whose previous loan was approved
- Refreshed clients who has unused loan offer
- Old female clients

Clients with Risk

- Lower Secondary educated clients are in most numbers to be defaulted when their previous loan were cancelled or refused

- Male clients with civil marriage

- Previously refused loan status group

Conclusion

- Driving variables which are strong indicators of default

    i. REGION_RATING_CLIENT
      - Selecting clients with highest region rating provided by bank itself will reduce risk of default
    ii. AMT_INCOME_TOTAL
    iii. OCCUPATION_TYPE
      - Selecting clients with higher occupation type will reduce the risk of default
    iv. NAME_EDUCATION_TYPE
      - Clients higher education and academic degree are less likely to default
    v. YEARS BIRTH
    vi. NAME_INCOME_TYPE